QTL linkage analysis in nuclear families

Peter M. Visscher
Queensland Institute of Medical Research
Brisbane, Australia
peter.visscher@qimr.edu.au
Overview

• QTL mapping by linear regression
  – Haseman-Elston regression
  – Using similarity scores

• Variance components & maximum likelihood

• IBD estimation from marker data

• Statistical power
Linkage = Co-segregation

Marker allele $A_1$ cosegregates with dominant disease
All QTL mapping methods are essentially 2-stage procedures

1. Genetic markers give information on IBD sharing between relatives [genotypes]

2. Association between phenotypes and genotypes gives information on QTL location and effect [linkage]

• Need informative mapping population
<table>
<thead>
<tr>
<th>Population</th>
<th>Features</th>
<th>Example Species</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inbred lines</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Backcross (BC)</td>
<td>Simplest design; powerful if dominance in ‘right’ direction</td>
<td>mice, plants</td>
</tr>
<tr>
<td>(F_2)</td>
<td>Estimation of additive and dominance effects; more powerful than BC for additive effects</td>
<td>mice, rats</td>
</tr>
<tr>
<td>Advanced intercross line (AIL)</td>
<td>As for (F_2) but with increased resolution of map location</td>
<td>mice</td>
</tr>
<tr>
<td>Recombinant inbred lines (RIL)</td>
<td>(F_1) followed by inbreeding; homozygous comparisons only; powerful for additive effects; less environmental noise</td>
<td>mice, plants</td>
</tr>
<tr>
<td>Congenic lines (= Nearly isogenic lines)</td>
<td>Backcrossing followed by inbreeding; homozygous comparisons only after inbreeding. Lines contain (\sim1%) of donor genome</td>
<td>mice, rats, plants</td>
</tr>
<tr>
<td>Double haploid lines (DHL)</td>
<td>Instant homzygosity through doubling of (F_1) gametes; homozygous comparisons only; powerful for additive effects and QTLxE interactions</td>
<td>plants</td>
</tr>
<tr>
<td>(F_{2:3})</td>
<td>Inbred progeny of (F_2); increased precision through progeny means</td>
<td>plants</td>
</tr>
<tr>
<td><strong>Structured outbred populations</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BC / (F_2) / AIL</td>
<td>As for inbred lines; mapping variation between lines</td>
<td>livestock, outbreeding trees/plants</td>
</tr>
<tr>
<td>Large fullsib families</td>
<td>Estimating contrasts between parental alleles. Allows for dominance estimation.</td>
<td>trees, fish, poultry</td>
</tr>
<tr>
<td>Halfsib families</td>
<td>Estimating contrasts between common parent alleles</td>
<td>cattle, pigs, poultry, trees</td>
</tr>
<tr>
<td>Nuclear families, including sibpairs</td>
<td>Detection of variance explained by markers</td>
<td>humans, livestock</td>
</tr>
<tr>
<td><strong>Unstructured outbred populations</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complex pedigrees</td>
<td>Detection of variance explained by markers</td>
<td>humans, livestock</td>
</tr>
</tbody>
</table>
Model: QTL as a random effect

\[ y_i = \mu + Q_i + A_i + E_i \]

\[ Q_i = \text{QTL genotype contribution for chrom. segment} \]

\[ A_i = \text{Contribution from rest of genome} \]

\[ \text{var}(y) = \sigma_q^2 + \sigma_a^2 + \sigma_e^2 \]
Genetic covariance between relatives

\[
\text{cov}(y_i, y_j) = \pi_{ij} \sigma_q^2 + a_{ij} \sigma_a^2
\]

\[a_{ij} = \text{average prop. of alleles shared in the genome (twice kinship coefficient)}\]

\[\pi_{ij} = \text{proportion of alleles IBD at QTL (0, } \frac{1}{2} \text{ or 1)}\]

\[E(\pi_{ij}) = a_{ij}\]
\[ \hat{\pi}_{ij} = \Pr(2 \text{ alleles IBD}) + \frac{1}{2}\Pr(1 \text{ allele IBD}) \]

\[ = \text{proportion of alleles IBD in non-inbred pedigree} \]

Estimate \( \hat{\pi}_{ij} \) with genetic markers
Random segregation and identity-by-descent in sibpairs
<table>
<thead>
<tr>
<th></th>
<th>Sib 1</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sib 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

|        |        |        |        |        |
|        |        |        |        |        |

4/16 = 1/4 sibs share BOTH parental alleles  IBD = 2
8/16 = 1/2 sibs share ONE parental allele  IBD = 1
4/16 = 1/4 sibs share NO parental alleles  IBD = 0
Several notations

<table>
<thead>
<tr>
<th>IBD</th>
<th>Probability</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBD0</td>
<td>$k_0$</td>
<td>0 or 1</td>
</tr>
<tr>
<td>IBD1</td>
<td>$k_1$</td>
<td>0 or 1</td>
</tr>
<tr>
<td>IBD2</td>
<td>$k_2$</td>
<td>0 or 1</td>
</tr>
<tr>
<td></td>
<td>$\Sigma=1$</td>
<td>$\Sigma=1$</td>
</tr>
</tbody>
</table>

\[\pi_a = \frac{1}{2}k_1 + k_2 = R = 2\theta\]
\[\pi_d = k_2 = \Delta_{xy}\]

Realisations

\[
\begin{array}{ccc}
k_0 & k_1 & k_2 \\
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
\end{array}
\]
### Sibpairs & fully informative marker

<table>
<thead>
<tr>
<th># Alleles</th>
<th>IBD</th>
<th>$\pi$</th>
<th>Pr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>$\frac{1}{4}$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>$\frac{1}{4}$</td>
</tr>
</tbody>
</table>

\[
E(\pi) = \Sigma \pi \Pr(\pi) = \frac{1}{2}
\]
\[
E(\pi^2) = \Sigma \pi^2 \Pr(\pi) = \frac{3}{8}
\]
\[
\text{var}(\pi) = E(\pi^2) - E(\pi)^2 = \frac{1}{8}
\]

\[
CV = 0.5\sqrt{2} = 70\%
\]
QTL mapping by linear regression

• Simple
  – standard stats package
• Robust
  – non-normal traits
• Powerful
• Computationally fast
  – permutations
  – bootstrapping
Haseman-Elston (1972)

“The more alleles pairs of relatives share at a QTL, the greater their phenotypic similarity”

or

“The more alleles they share IBD, the smaller the difference in their phenotype”
Population sib-pair trait distribution
No linkage
Under linkage
Sib pair design to map QTL

• Multiple ‘families’ of two (or more) sibs
• Phenotypes on sibs
• Marker genotypes on sibs (& parents)

• Correlate phenotypes and genotypes of sibs
Data structure is simple

<table>
<thead>
<tr>
<th>Pair</th>
<th>Phenotypes</th>
<th>Prop. alleles IBD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$y_{11}$</td>
<td>$\pi_1$</td>
</tr>
<tr>
<td>2</td>
<td>$y_{21}$</td>
<td>$\pi_2$</td>
</tr>
<tr>
<td>2</td>
<td>$y_{22}$</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>$y_{n1}$</td>
<td>$\pi_n$</td>
</tr>
<tr>
<td>n</td>
<td>$y_{n2}$</td>
<td></td>
</tr>
</tbody>
</table>

$\pi = 0, \frac{1}{2} \text{ or } 1 \text{ for fully informative markers}$
Properties of squared differences

\[ E(Y_1 - Y_2)^2 = \text{var}(Y_1 - Y_2) + (E(Y_1 - Y_2))^2 \]

\[ \text{var}(Y_1 - Y_2) = \text{var}(Y_1) + \text{var}(Y_2) -2\text{cov}(Y_1,Y_2) \]

If \( E(Y_i) = E(Y_j) \) and \( \text{var}(Y_1)=\text{var}(Y_2) \), then

\[ E(Y_1 - Y_2)^2 = 2(1 - \rho)\text{var}(Y) \text{ and } \]
\[ \text{Var}(Y_1 - Y_2)^2 = 8(1 - \rho)^2 \text{ var}(Y)^2 \]

If \( Y \sim \text{N}[E(Y),\text{var}(Y)] \)}
Haseman-Elston method

- Phenotypes on a relative pair:

\[ Y = (y_1 - y_2)^2 \]

\[ \mathbb{E}(Y) = \mathbb{E}[(Q_1 - Q_2)^2 + (A_1 - A_2)^2 + (E_1 - E_2)^2] \]

\[ = \mathbb{E}[(Q_1 - Q_2)^2] + \{2(1-a_{12})\sigma_a^2 + 2\sigma_e^2\} \]

\[ = 2[\sigma_q^2 - \text{cov}(Q_1,Q_2)] + \{\sigma_e^2\} \]

\[ = (2\sigma_q^2 + \sigma_e^2) - 2\pi_{12}\sigma_q^2 \]

\[ \pi_{12} = \text{proportion of alleles IBD at QTL for the pair of relatives} \]
Conditional expectation

\[ E(Y_j \mid \pi_j) = (2\sigma_q^2 + \sigma_\varepsilon^2) - (2\sigma_q^2)\pi_j \]

- negative slope of \( Y \) on \( \pi \) if \( \sigma_q^2 > 0 \)
- estimate \( \pi_j \) from marker data \([ \hat{\pi}_j \])
- use simple linear regression to detect QTL:

\[ E(Y_j \mid \hat{\pi}_j) = \alpha + \beta\hat{\pi}_j \]
A significant negative slope indicates linkage to a QTL.
Single fully informative marker

\[ \beta = -2(1 - 2r)^2 \sigma_q^2 \]
\[ \alpha = 2[1 - 2(1-r)r] \sigma_q^2 + \sigma_\varepsilon^2 \]
\[ r = \text{recombination fraction between marker & QTL} \]

- Disadvantage of method
  - not powerful
  - confounding between QTL location and effect

- Use multipoint estimates of \( \pi \)
Example from Cardon et al. (1994)

Be sceptical if you see plots like this!
Squared difference (again)

\[ E(Y_1 - Y_2)^2 = 2(1-\rho)\text{var}(Y) \]
\[ \text{Var}(Y_1 - Y_2)^2 = 8(1 - \rho)^2 \text{var}(Y)^2 \]
\[ \text{SD}(Y_1 - Y_2)^2 = (2\sqrt{2})(1 - \rho) \text{var}(Y) \]

NB: Linkage Practical
Distribution of test statistic under the null hypothesis of no linkage

<table>
<thead>
<tr>
<th>Method</th>
<th>Null</th>
<th>Alternative</th>
<th>Action if wrong alternative</th>
</tr>
</thead>
<tbody>
<tr>
<td>HE</td>
<td>$\beta=0$</td>
<td>$\beta&lt;0$</td>
<td>Set $b=0$</td>
</tr>
<tr>
<td>VC</td>
<td>$\sigma_{QTL}^2=0$</td>
<td>$\sigma_{QTL}^2&gt;0$</td>
<td>Set $\sigma_{QTL}^2=0$</td>
</tr>
</tbody>
</table>

- The statistical tests are one-sided.
  - If Null is true then expect $b>0$ or $\sigma_{QTL}^2<0$ with a probability of $\frac{1}{2}$
  - The test statistic follows a mixture of distributions with mixing proportion of $\frac{1}{2}$
  - LRT $\sim 0$ with prob $= \frac{1}{2}$ and $\sim \chi^2_{(1)}$ with prob $= \frac{1}{2}$
  - For data analysis: take p-values from $\chi^2_{(1)}$ and divide by two
QTL mapping using similarity scores

• Sometimes we only know (dis)similarity between pairs of relatives
  – mosquito bites
  – eye colour

• H-E principle still applies: more similar implies more alleles shared IBD
Models
(d= dissimilarity score; s= similarity score)

\[ d = \mu_d + \beta_d \pi \]

\[ H_1: \beta_d < 0 \]

\[ s = \mu_s + \beta_s \pi \]

\[ H_1: \beta_s > 0 \]
Replicated Linkage for Eye Color on 15q Using Comparative Ratings of Sibling Pairs

Danielle Posthuma,1,4 Peter M. Visscher,2 Gonneke Willemsen,1 Gu Zhu,2 Nicholas G. Martin,2 P. Eline Slagboom,3 Eco J. C. de Geus,1 and Dorret I. Boomsma1

Eye color similarity was rated on a three point scale ("not at all alike"—"somewhat alike"—"completely alike"). The probability that twins were alike for eye color (eye color similarity, s) was calculated from the response pattern on all questionnaires and all informants, by summing over the product of the three possible answer categories (where "not at all alike" was coded 0, "somewhat alike" was coded 0.5, and "completely alike" was coded 1) and their
Fig. 3. Chromosome 15 region of significant linkage for eye color. The triangle marks the location of the linkage peak from Zhu et al. (2004) on 15q. Dotted lines represent the positions of the markers. The x-axis is in centiMorgan.
Linkage by Variance Component Analysis

• Why?
  – More powerful (in theory)
  – Logical extension of analysis of resemblance between relatives
  – Applicable to general (& large) pedigrees
Variance-Covariance Matrix for a pair of relatives

\[ \Sigma = \begin{bmatrix} V(y_1) & \text{Cov}(y_1, y_2) \\ \text{Cov}(y_1, y_2) & V(y_2) \end{bmatrix} \]

Model must describe not only variance of each observation but also covariance for pairs of observations.
Variance-Covariance Matrix

\[
\Sigma_{jk} = \begin{cases} 
\sigma_q^2 + \sigma_a^2 + \sigma_c^2 + \sigma_e^2 & \text{if } j = k \\
\hat{\pi}_{jk} \sigma_q^2 + a_{jk} \sigma_a^2 + \sigma_c^2 & \text{if } j \neq k 
\end{cases}
\]

\(a_{jk}\) is twice the kinship coefficient,

= i.e. twice the probability that two genes sampled at random from a pair of individuals are identical. (1 for MZ twins and 0.5 for fullsibs)
Variance-Covariance Matrix

\[ \Sigma_{jk} = \begin{cases} 
\sigma_q^2 + \sigma_a^2 + \sigma_c^2 + \sigma_e^2 & \text{if } j = k \\
\hat{\pi}\sigma_q^2 + \rho\sigma_a^2 + \sigma_c^2 & \text{if } j \neq k 
\end{cases} \]

\( \sigma^2 = \) variation due to:
\( q = \) QTL;
\( a = \) polygenic;
\( e = \) individual-specific
\( c = \) shared environment
Bivariate density function

• Normal density function

\[ L(y) = \frac{1}{\sqrt{2\pi}} \sigma^{-1} e^{-\frac{1}{2}(y-\mu)^2 / \sigma^2} \]

• Bivariate normal density function

\[ L(y) = \frac{1}{2\pi} \left| \Sigma \right|^{-1/2} e^{-\frac{1}{2}(y-\mu)'\Sigma^{-1}(y-\mu)} \]
Alternate hypothesis of linkage for sibpairs (Likelihood function):

\[
L_i = \prod_i \left(2\pi\right)^{-1} \left| \Sigma_i \right|^{-\frac{1}{2}} e^{-\frac{1}{2}(y_i - \mu)^T \Sigma^{-1} (y_i - \mu)}
\]

\[
\Sigma = \begin{bmatrix}
\hat{\pi}_q \sigma_q^2 + \sigma_a^2 + \sigma_c^2 + \sigma_e^2 \\
\pi_i \sigma_q^2 + a_i \sigma_a^2 + \sigma_c^2 \\
\sigma_q^2 + \sigma_a^2 + \sigma_c^2 + \sigma_e^2
\end{bmatrix}
\]

Null hypothesis:

\[
L_0 = \prod_i \left(2\pi\right)^{-1} \left| \Sigma_i \right|^{-\frac{1}{2}} e^{-\frac{1}{2}(y_i - \mu)^T \Sigma^{-1} (y_i - \mu)}
\]

\[
\Sigma = \begin{bmatrix}
\sigma_a^2 + \sigma_c^2 + \sigma_e^2 \\
\sigma_a^2 + \sigma_c^2 \\
\sigma_a^2 + \sigma_c^2 + \sigma_e^2
\end{bmatrix}
\]

Note three uses of 'pi'!
Test statistic ML

\[ LRT = 2\ln(ML_{\text{full}}) - 2\ln(ML_{\text{reduced}}) \]

\[ H_0: LRT \sim \frac{1}{2} \chi^2(1) + \frac{1}{2}(0) \]
IBD calculating algorithms

- **Elston-Stewart algorithm**
  Handles large pedigrees, but small nr of loci, exact IBD distributions (Elston and Stewart, 1971)

- **Lander-Green algorithm**
  Handles small pedigrees, but large nr of loci, exact IBD distributions (Lander and Green, 1987)

- **MCMC methods**
  Calculates approximate IBD distributions (Heath, 1997)

- **Average sharing methods**
  Calculates approximate IBD distributions (Fulker et al., 1995; Almasy and Blangero, 1998)
Estimating $\pi$ when marker is not fully informative

• Using:
  – Mendelian segregation rules
  – Marker allele frequencies in the population
IBD can be trivial...

IBD=0
Two Other Simple Cases…

IBD=2
A little more complicated...

IBD=1 (50% chance)

IBD=2 (50% chance)
And even more complicated...

IBD=?  1/1  1/1
Bayes Theorem for IBD Probabilities

\[
P(\text{IBD} = i \mid G) = \frac{P(\text{IBD} = i, G)}{P(G)}
\]

\[
= \frac{P(\text{IBD} = i)P(G \mid IBD = i)}{P(G)}
\]

\[
= \frac{\sum_{j} P(\text{IBD} = j)P(G \mid IBD = j)}{\sum_{j} P(\text{IBD} = j)P(G \mid IBD = j)}
\]
### P(Marker Genotype | IBD State)

<table>
<thead>
<tr>
<th>Sib</th>
<th>CoSib</th>
<th>IBD</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(a,b)</td>
<td>(c,d)</td>
<td>0</td>
<td>$p_ap_bp_c p_d$</td>
</tr>
<tr>
<td>(a,a)</td>
<td>(b,c)</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>(a,a)</td>
<td>(b,b)</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>(a,b)</td>
<td>(a,c)</td>
<td>0</td>
<td>$p_a p_bp_c$</td>
</tr>
<tr>
<td>(a,a)</td>
<td>(a,b)</td>
<td>0</td>
<td>$p_a^2 p_b$</td>
</tr>
<tr>
<td>(a,b)</td>
<td>(a,b)</td>
<td>0</td>
<td>$p_a p_b^2 + p_a^2 p_b$</td>
</tr>
<tr>
<td>(a,a)</td>
<td>(a,a)</td>
<td>0</td>
<td>$p_a$</td>
</tr>
</tbody>
</table>

**Prior Probability**

- $\frac{1}{4}$
- $\frac{1}{2}$
- $\frac{1}{4}$

[Assumes Hardy-Weinberg proportions of genotypes in the population]
$p_1 = 0.5$

$P(G | IBD = 0) = p_1^4 = \frac{1}{16}$

$P(G | IBD = 1) = p_1^3 = \frac{1}{8}$

$P(G | IBD = 2) = p_1^2 = \frac{1}{4}$

$P(G) = \frac{1}{4} p_1^4 + \frac{1}{2} p_1^3 + \frac{1}{4} p_1^2 = \frac{9}{64}$

$P(IBD = 0 | G) = \frac{\frac{1}{4} p_1^4}{P(G)} = \frac{1}{9}$

$P(IBD = 1 | G) = \frac{\frac{1}{2} p_1^3}{P(G)} = \frac{4}{9}$

$P(IBD = 2 | G) = \frac{\frac{1}{4} p_1^2}{P(G)} = \frac{4}{9}$

$\hat{\pi} = \frac{2}{3}$
Statistical power

• Calculate the expected value of the test statistic under the alternative hypothesis
• Calculate the variance of the test statistic under the alternative hypothesis
→ Distribution of test statistic under null and alternative hypotheses → power of detection
Linear regression (HE)

\[ D = (y_1 - y_2) \]
\[ D^2 = (y_1 - y_2)^2 \]
Regression

\[ Y = \mu + \beta \pi + e \]

Test statistic \[ = \frac{\hat{\beta}^2}{\text{var}(\hat{\beta})} \]
\[ \sim \text{(non)central } \chi^2 \]

If we know \( \beta \), \( \text{var}(Y) \) and \( \text{var}(\pi) \), we can predict the expected test statistic & power
$\beta$ for additive QTL model (assume $\sigma_y = 1$)

$\beta = -2q^2$

With $q^2$ the proportion of phenotypic variation due to the QTL
var(\hat{\beta}) for additive model

\[ Y = \mu + \beta \pi + e \]

\[ \text{var}(\hat{\beta}) \approx \frac{\text{var}(e)}{[(n-2)\text{var}(\pi)]} \]

\[ \text{var}(e) \approx \text{var}(Y) - \beta^2 \text{var}(\pi) \]
Analytical predictions

\[
\beta = -2q^2
\]

\[
\text{var}(\pi) = \frac{1}{8}
\]

\[
\text{var}(Y) = \text{Depends on population parameters;}
\]

\[
\text{var}(y) = f^2 + q^2 + r^2
\]

- Variance due to family effects
- Variance due to QTL
- Residual variance
QTL models

- Random QTL effects
  \[ Q_1 \sim N(0, q^2) \]
  \[ Q_2 | \pi \sim N(\pi Q_1, (1 - \pi^2)q^2) \]

- Fixed QTL effect
  - bi-allelic additive QTL with frequency \( p \)
  - \( q^2 = 2p(1-p)a^2 \)
Variance of Y for random QTL (exact)

\[ \text{var}(D^2) = 8r^2(1 - f^2) + (7/2)q^4 \]

[Visscher & Hopper 2001, Annals Human Genetics]
If QTL effects is small (approximation)

- Bivariate normality, sib-correlation $\rho$

$$\text{var}(D^2) \sim 8(1 - \rho)^2$$

$$\rho = f^2 + \frac{1}{2}q^2$$
Haseman-Elston regression

\[
E(\text{test statistic} \mid \text{QTL}) = E[\hat{\beta}^2 / \text{var}(\hat{\beta})] \\
= \frac{1}{2} + \frac{1}{2}nq^4 / \text{var}(D^2) \\
E(T) = \frac{1}{2} + n\lambda \\
\lambda = q^4 / [16(1-\rho)^2]
\]
Sample size for QTL detection in genome scans using sibpairs

Power = 90%. Type-I error = $10^{-5}$
Maximum likelihood
(assuming bivariate normality | $\pi$)

Full model:

$$-2\ln(L) = \sum n_{\pi} \ln|V_{\pi}| + \sum (y-\mu)'V_{\pi}^{-1}(y-\mu)$$

$$V_{\pi} = \begin{pmatrix} f^2 + q^2 + r^2 & f^2 + \pi q^2 \\ f^2 + \pi q^2 & f^2 + q^2 + r^2 \end{pmatrix}$$
Maximum likelihood

Reduced model:

\[-2\ln(L) = n\ln|V| + (y-\mu)'V^{-1}(y-\mu)\]

\[V = \begin{pmatrix} f^2 + r^2 & f^2 \\ f^2 & f^2 + r^2 \end{pmatrix}\]

[Fulker & Cherny 1996; Wright 1997]
Maximum Likelihood

$$\text{LRT} = 2\ln(\text{ML}_{\text{full}}) - 2\ln(\text{ML}_{\text{reduced}})$$

$$E(\text{LRT}|\text{QTL}) \approx \frac{1}{2} + n\lambda$$

$$\lambda = \ln\left\{\left[1-(f^2+\frac{1}{2}q^2)^2\right]^{1/2} / \left[(1-f^4)(1-(f^2+q^2)^2\right]^{1/4}\right\}$$

$$\sim q^4/8 \text{ if all correlations are small}$$

[Fulker & Cherny 1996; Wright 1997]
Genetic Power Calculator (PGC)
http://pngu.mgh.harvard.edu/~purcell/gpc/

Genetic Power Calculator
S. Purcell & P. Sham, 2001-2009

This site provides automated power analysis for variance components (VC) quantitative trait locus (QTL) linkage and association tests in sibships, and other common tests. Suggestions, comments, etc to Sham Purcell.

If you use this site, please reference the following Bioinformatics article:


Genetic Power Calculator

Modules

- Case-control for discrete traits
- Case-control for threshold-selected quantitative traits
- QTL association for sibships and singletons
- TDT for discrete traits
- TDT and parentTDT with ascertainment
- TDT for threshold-selected quantitative traits
- Epistasis power calculator
- QTL linkage for sibships
- Probability Function Calculator

QTL Linkage for Sibships

- QTL additive variance
- QTL dominance variance
- Residual shared variance
- Residual nonshared variance
- Recombination fraction

Sample Size

Sibship Size

User-defined type I error rate

User-defined power: determine N

(1 - type II error rate)
LOD score and likelihood-ratio test statistic (LRT)

\[ \text{LRT} = 2\ln(L_1/L_0) = 2[\ln(L_1)-\ln(L_0)] = -2[\ln(L_0)-\ln(L_1)] \]

\[
\text{LOD} = \log_{10}(L_1/L_0) = \frac{\ln(L_1/L_0)}{\ln(10)} = \frac{2\ln(L_1/L_0)}{2\ln(10)} = \frac{\text{LRT}}{4.605} = 0.217 \text{ LRT}
\]
# Test statistics for linkage analysis

<table>
<thead>
<tr>
<th>Method</th>
<th>Test</th>
<th>Asymptotic Distr.</th>
</tr>
</thead>
</table>
| Linear Regression             | t-test or F-test            | $t_n \approx \sim N(0,1)$  
|                               |                             | $F_{k,n} \approx \sim (1/k)\chi^2_{(k)}$ |
| Maximum likelihood            | LRT = $2\ln(L_1/L_0)$       | $\sim \chi^2_{(k)}$ |
| Maximum likelihood            | LOD = $\log_{10}(L_1/L_0)$  | $\sim 0.217\chi^2_{(k)}$ |
| -log_{10}(p-value)            | Arbitrary                   | -2ln(p) $\sim \chi^2_{(2)}$ |
| Non-parametric                | Z-score                     | $Z \sim N(0,1)$  
|                               |                             | $Z^2 \sim \chi^2_{(1)}$ |
Software for linkage analyses

- Genehunter
- Mendel
- Vitesse
- Allegro
- Simwalk
- Loki
- Merlin (Computer Practical)
- Solar